IAN FOSTER INTERVIEW

How is grid computing today similar to the Internet in the early nineties?

In the early nineties, the Internet made a transition from being something of utility, and really only known to people in academia, to something that was of broad industrial relevance with the emergence of the Mosaic web browser and Netscape. Grid computing is making a similar transition at the moment, from academia to industry. Just in the last two years or so we've seen major corporations, like IBM, Sun, and HP deploying grid products.

What is the history of distributed computing?

Grid computing is about the large-scale integration of computing systems to enable new classes of applications to provide on-demand access to computing and information. And that's certainly not a new idea. Back in 1969, I think, when the very first node of the (then) ARPANET, which became the Internet, was deployed at UCLA in Los Angeles, the people there put out a press release touting the wonderful things that were going to happen once the Internet was ubiquitous. It was quite an ambitious and visionary view of things given that they only had one node at that point! What's different now is that we have the quasi-ubiquitous Internet networks suddenly getting fast enough that we can connect our computers and people and information sources in ways that were not possible before. Also, the software has evolved to the point where we can start thinking about linking distributed computing systems into something really interesting.

When did our computer processing capacity and data storage capacity reach the point where we could even consider implementing distributed computing utilities or grids?

One important part of the evolution towards grid computing is, in addition to the deployment of the Internet, the fact that our home computers are now as powerful as yesterday's supercomputers. As we all know, the power of our computers continues to double every eighteen months or so. The laptop that I use for most of my work nowadays, for example, is faster than the supercomputers that were deployed at the U.S. national centers just ten years ago. This process is an ongoing one and the same observation will probably be true ten years from now. But the transition to a system that enables participation in a grid as a true peer is something that's just happened in the last five years or so.

What was it like back in the day when you and your colleague Dr. Kesselman envisioned designing software that could juggle and link all the computing resources across sites and deliver them on demand?

Our thinking on these topics, of course, has evolved over the years, but I think the vision has stayed fairly consistent. We started on this in the mid-nineties when the high-speed networks were starting to be deployed. We had a strong interest in understanding how computers, and distributed systems for that matter, could be used to enhance the process of scientific discovery. And we realized that there was a need for a new class of technology that would allow people to build distributed computing systems more easily than was then the case.

So we put together a small team, coined the name Globus, and we started working on some basic software with scientific teams who were eager at the time to try new technologies. It has gradually scaled up from a few partners to an increasing number of large scientific projects and, more recently, industrial groups.

What is the difference between grid computing and distributed computing schemes such as SETI@Home?

SETI@Home is a simple and very effective example of grid computing. It is one of the larger examples of a number of systems that sort of behave in similar ways, whose goal is to harness idle computers in order to perform a large number of computational tasks that are essentially identical but different in terms of the data that is being processed. In the case of SETI@Home, the problem is that of searching for extraterrestrial intelligence by processing data from the Arecibo radio telescope. Every home computer that signs up receives periodically a piece of radio telescope data and processes it using signal processing algorithms that look for particular signatures that correspond to a signal. With SETI@Home, the goal is indeed large-scale integration of computing systems, but the actual communication patterns that are going on are pretty simple. Grid technology involves far more complex couplings of systems at different locations, and often much larger volumes of data and more challenging computational tasks.

What is the TeraGrid?

The TeraGrid is one of the most ambitious scientific grid projects underway. This is a U.S. project funded by the National Science Foundation. It involves the construction of very large computer systems at four sites across the U.S. – two in Illinois and two in California. The linking of those systems by a very high-speed network, one that runs at 40 gigabytes per second, which is a million times faster than a typical home modem. These four sites and a number of others are integrated into a grid system so that a scientist with access can sign on and obtain on-demand access to any one of the systems in this grid infrastructure.

So it's the linking of supercomputers?

Yes. This is one important application of grid computing but certainly not the only one. There's an interesting project called Grid2003 that links more medium scale computer clusters at thirty sites or so across the U.S., and across to computers in Europe as well. Grid2003 is intended for large-scale data analysis in the physics community.

How is NASA tapping into the grid?

NASA is an early adopter. They created a project some years ago called the NASA Information Power Grid (IPG) which aims to link together some of the major NASA laboratories. Some of their goals involve the design of aerospace vehicles; they're interested in allowing distributed teams to formulate and then run numerical simulations of these systems. One of the challenges they face is the number of different components involved: there could be an airframe model, an airflow model and the combustion model for the engine, each developed by different groups. Grid computing links them together to build a numerical simulation of the ensemble.

How does the Globus Toolkit software change the way earthquake engineers

work?

Earthquake engineers are concerned with the design of earthquake-resistant buildings. And one of the resources they use for that work are shake tables, very large structures on which you put model buildings and simulate the effects of earthquakes.
The engineers used to have to travel great distances to facilities to construct and run experiments, and the results were typically fairly closely held. So the National Science Foundation in the U.S. has funded the Network for Earthquake Engineering Simulation (NEESgrid), which links together the earthquake engineering facilities in the U.S. with their remote user community, and with various data archives and simulation and computers, to form what's called a "collaboratory," which is another word for grid, if you like.

What grid developments are you most excited about?

I remain delighted by the range of applications in the sciences, as with the earthquake engineers, the astronomers, and more recently the biologists. I think there are tremendous opportunities to accelerate work that will uncover the secret of life at a very fundamental level. It's also wonderful seeing what's going on in industry. In the next few years, I believe we'll realize this vision of a universal computing facility.

Ian Foster is Associate Director in the Mathematics and Computer Science Division at Argonne National Laboratory and the Arthur Holly Compton professor of computer science at the University of Chicago.

Massive Change Radio is a project by Jennifer Leonard, Bruce Mau Design, the Institute without Boundaries, and CIUT FM, Toronto.